The Computation of Previously Inaccessible Digits of π^2 and Catalan's Constant

David H. Bailey[*]   Jonathan M. Borwein[†]   Andrew Mattingly[‡]

Glenn Wightwick[§]

April 19, 2011

1   Introduction

The poem "Pi" by 1996 Nobel Laureate Wislawa Szymborska starts:

> The admirable number pi:
> *three point one four one.*
> All the following digits are also initial,
> *five nine two* because it never ends.
> It can't be comprehended *six five three five* at a glance,
> *eight nine* by calculation,
> *seven nine* or imagination,
> not even *three two three eight* by wit, that is, by comparison
> *four six* to anything else
> *two six four three* in the world.
> The longest snake on earth calls it quits at about forty feet.
> Likewise, snakes of myth and legend, though they may hold out a bit longer.
> The pageant of digits comprising the number pi
> doesn't stop at the page's edge.

We have recently concluded a very large mathematical calculation, uncovering objects that until recently were widely considered to be forever inaccessible to computation. These calculations required both human ingenuity and the extraordinary power of modern highly parallel computer technology.

Our computations stem from the "BBP" formula for π, which was discovered in 1997 using a computer program implementing the "PSLQ" integer relation algorithm. This formula has

the remarkable property that it permits one to directly calculate binary digits of π, beginning at an arbitrary position d, without needing to calculate any of the first $d-1$ digits. Since 1997, numerous other BBP-type formulas have been discovered for various mathematical constants, including formulas for π^2 (both in binary and ternary bases), and for Catalan's constant.

In this article we describe the computation of base-64 digits of π^2, base-729 digits of π^2, and base-4096 digits of Catalan's constant, in each case beginning at the ten trillionth place, computations that involved a total of approximately 1.549×10^{19} floating-point operations. We also discuss connections between BBP-type formulas and the age-old unsolved questions of whether and why constants such as $\pi, \pi^2, \log 2$ and Catalan's constant have "random" digits.

2   A brief history of the computation of pi

Since the dawn of civilization, mathematicians have been intrigued by the digits of π [11]. In the third century BCE, Archimedes employed a brilliant scheme of inscribed and circumscribed $3 \cdot 2^n$-gons to compute π to two decimal digit accuracy [8]. It was the first true algorithm for π, in that it permitted one to produce an arbitrarily accurate value for π. However, this and other numerical calculations of antiquity were severely hobbled by their reliance on primitive arithmetic systems.

A major breakthrough in this regard, which some regard as among the greatest scientific developments of all time, was the discovery of full positional decimal arithmetic with zero, by an unknown mathematician in fifth century India. Several hundred years later, in 999 CE, scientist-Pope Sylvester II attempted to introduce decimal arithmetic in Europe. Yet, little headway was made until the publication of Fibonacci's *Liber Abaci* in 1202, and several hundred more years would pass before the system finally gained universal, if belated, adoption in the West. The time of Sylvester's reign was a very turbulent one, and he died in 1002, shortly after the death of his protector Emperor Otto III. It is interesting to speculate how history would have changed had he lived longer. A page from his mathematical treatise *De Geometria* is shown in Figure 1.

2.1   Pi after calculus

Armed with decimal arithmetic, and spurred by the newly discovered methods of calculus, mathematicians computed π with aplomb. Isaac Newton himself devised an arcsine-like scheme to compute digits of π and recorded 15 digits, although he sheepishly acknowledged, "I am ashamed to tell you to how many figures I carried these computations, having no other business at the time." Newton wrote these words during the plague year 1666, when, ensconced in a country estate, he devised the fundamentals of calculus and the laws of motion and gravitation.

All records until 1980 relied on so called *Machin-type formulas* [10], which write π as a linear combination of arctangent values. The most famous of these is Machin's formula:

$$\frac{\pi}{4} = 4\arctan\left(\frac{1}{5}\right) - \arctan\left(\frac{1}{239}\right). \tag{1}$$

In 1844 the Viennese *computer* and *kopfrechner* Johan Zacharias Dase (1824 -1861) calcu-
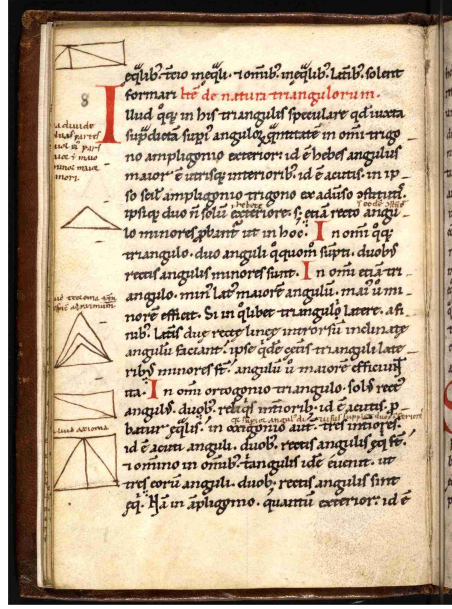
Figure 1: Excerpt from *de Geometria* by Pope Sylvester II (reigned 999-1002 CE)

lated π to 200 places upon learning Euler's Machin-type formula

$$\frac{\pi}{4} = \tan^{-1}\left(\frac{1}{2}\right) + \tan^{-1}\left(\frac{1}{5}\right) + \tan^{-1}\left(\frac{1}{8}\right) \tag{2}$$

from Strassnitsky. He did this feat in his head, correctly, over a two month period, without any aids.

The culmination of these feats was a computation of π using (1) to 527 digits in 1853 by William Shanks, later (erroneously) extended to 707 digits — see Table 1. In the preface to the publication of this computation, Shanks wrote that his work "would add little or nothing to his fame as a Mathematician, though it might as a Computer" (until 1950 a "computer" was a person, and a "calculator" was a machine).

One motivation for computing these digits was to see whether the digits of π repeat, thus disclosing the fact that π is a ratio of two integers. In 1766, Lambert proved that π is irrational, by means of a continued fraction argument, thus establishing that the digits of π do not repeat in any number base. In 1882, Lindemann established that π is transcendental, thus establishing that the digits of π^2 or any integer polynomial of π cannot repeat.

2.2 Pi in the computer age

At the dawn of the computer age, John von Neumann suggested computing digits of π and e for statistical analysis, and in 1949 π was computed to 2037 digits, at the instigation of John von Neumann, on the *Electronic Numerical Integrator And Calculator* (ENIAC) — see Figure 2. In 1965 mathematicians realized that the newly-discovered fast Fourier transform could be used to dramatically accelerate high-precision multiplication.

In 1976, Eugene Salamin and Richard Brent independently discovered a new algorithm for π based on elliptic integrals and the Gauss arithmetic-geometric mean iteration [10]. This

Name	Year	Digits
Babylonians	2000? BCE	1
Egyptians	2000? BCE	1
Hebrews (1 Kings 7:23)	550? BCE	1
Archimedes	250? BCE	3
Ptolemy	150	3
Liu Hui	263	5
Tsu Ch'ung Chi	480?	7
Madhava	1400?	13
Al-Kashi	1429	14
Romanus	1593	15
Van Ceulen	1615	35
Sharp (and Halley)	1699	71
Machin	1706	100
Strassnitzky and Dase	1844	200
Rutherford	1853	440
W. Shanks	1853	(607) 527
W. Shanks	1874	(707) 527
Ferguson (Calculator)	1947	808

Table 1: Pre-computer-age π computations

Figure 2: The ENIAC in the Smithsonian Museum

Name	Year	Correct Digits
Reitwiesner et al. (ENIAC)	1949	2,037
Genuys (IBM)	1958	10,000
D. Shanks and Wrench (IBM)	1961	100,265
Guilloud and Bouyer (IBM)	1973	1,001,250
Miyoshi and Kanada	1981	2,000,036

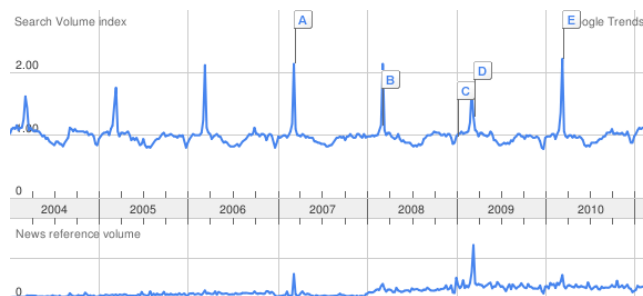Table 2: Early computer-era π calculations

Figure 3: Searches for "Pi" on Google Trends, showing spikes near the annual Pi Day (3.14).

converges much more rapidly than the classical infinite series used since the time of Newton—each iteration approximately *doubles* the number of correct digits in the result. When these and other algorithmic advances were combined with the exponentially increasing power of computer hardware, π was computed to over one million digits in 1973, to over one billion digits in 1989, and to over one trillion digits in 2002 — see Tables 2 and 3.

As recently as 1963, Daniel Shanks, who himself calculated π to over 100,000 digits, told Philip Davis that computing one billion digits would be "forever impossible." Yet this feat was achieved less than 30 years later in 1989 by Yasumasa Kanada of Japan. Also, in 1989, famous British physicist Roger Penrose, in the first edition of his best-selling book *The Emperor's New Mind*, declared that humankind likely will never know if a string of ten consecutive sevens occurs in the decimal expansion of π. This string was found just eight years later, in 1997, also by Kanada, beginning at position 22,869,046,249. After being advised by one of the present authors, Penrose revised his second edition to specify twenty consecutive sevens.

Kanada also found the string 0123456789, which plays a famous role in the intuitionist philosophy of mathematics, commencing at the 17,387,594,880-th position after the decimal point. This was despite Brouwer and Heyting's certainty that when and if it occurred was unknowable. Even astronomer Carl Sagan, whose lead character in his 1985 novel *Contact* (played by Jodi Foster in the movie version) sought confirmation in base-11 digits of π, expressed surprise to learn, shortly after the book's publication, that π had already been computed to many millions of digits.

So much for human certainty of continued human ignorance. Jonathan and Peter Borwein, when asked about π in 1986 by the *Los Angeles Times*, opted to be correct in their lifetimes and replied that 10^{100} digits was out of sight. They have not yet found it necessary to retract.

In spite of these advances and many additional mathematical discoveries [8], an air of mystery still surrounds π. As a single example, the question of whether, or why, the digits of π appear statistically random remains completely unanswered, a glaring hole in mathematical knowledge that is laid bare whenever a child sees the digits of π printed on a banner in her classroom and asks "Why is there no pattern?"

The one known regular pattern regarding π is in the annual search for information around Pi Day (March 14, i.e., 3/14), as Figure 3, taken from `http://www.google.com/trends?q=Pi+`, makes clear.

5

Name	Year	Correct Digits
Miyoshi and Kanada	1981	2,000,036
Kanada-Yoshino-Tamura	1982	16,777,206
Gosper	1985	17,526,200
Bailey	Jan. 1986	29,360,111
Kanada and Tamura	Sep. 1986	33,554,414
Kanada and Tamura	Oct. 1986	67,108,839
Kanada et. al	Jan. 1987	134,217,700
Kanada and Tamura	Jan. 1988	201,326,551
Chudnovskys	May 1989	480,000,000
Kanada and Tamura	Jul. 1989	536,870,898
Kanada and Tamura	Nov. 1989	1,073,741,799
Chudnovskys	Aug. 1991	2,260,000,000
Chudnovskys	May 1994	4,044,000,000
Kanada and Takahashi	Oct. 1995	6,442,450,938
Kanada and Takahashi	Jul. 1997	51,539,600,000
Kanada and Takahashi	Sep. 1999	206,158,430,000
Kanada-Ushiro-Kuroda	Dec. 2002	1,241,100,000,000
Takahashi	Jan. 2009	1,649,000,000,000
Takahashi	Apr. 2009	2,576,980,377,524
Bellard	Dec. 2009	2,699,999,990,000
Kondo and Yee	Aug. 2010	5,000,000,000,000

Table 3: Modern computer-era π calculations

3 The BBP formula for pi

A 1997 paper [3], [9, Ch. 3] by one of the present authors (Bailey), Peter Borwein (brother of Jonathan Borwein) and Simon Plouffe presented the following previously unknown formula for π, now known as the "BBP" formula for π:

$$\pi = \sum_{k=0}^{\infty} \frac{1}{16^k} \left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right). \tag{3}$$

This formula has the remarkable property that it permits one to directly calculate binary or hexadecimal digits of π beginning at an arbitrary starting position, without needing to calculate any of the preceding digits. The resulting simple algorithm requires only minimal memory, does not require multiple-precision arithmetic, and is very well suited to highly parallel computation. The cost of this scheme increases only slightly faster than the index of the starting position.

The proof of this formula is surprisingly elementary. First note that for any $k < 8$,

$$\int_0^{1/\sqrt{2}} \frac{x^{k-1}}{1-x^8} \, \mathrm{d}x \; = \; \int_0^{1/\sqrt{2}} \sum_{i=0}^{\infty} x^{k-1+8i} \, \mathrm{d}x \; = \; \frac{1}{2^{k/2}} \sum_{i=0}^{\infty} \frac{1}{16^i (8i+k)}. \tag{4}$$

Thus one can write

$$\sum_{i=0}^{\infty} \frac{1}{16^i} \left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6} \right) = \int_0^{1/\sqrt{2}} \frac{4\sqrt{2} - 8x^3 - 4\sqrt{2}x^4 - 8x^5}{1-x^8} \, \mathrm{d}x, \tag{5}$$

which on substituting $y := \sqrt{2}x$ becomes

$$\int_0^1 \frac{16\,y - 16}{y^4 - 2\,y^3 + 4\,y - 4}\, \mathrm{d}y \;=\; \int_0^1 \frac{4y}{y^2 - 2}\, \mathrm{d}y - \int_0^1 \frac{4y - 8}{y^2 - 2y + 2}\, \mathrm{d}y \;=\; \pi, \tag{6}$$

reflecting a partial fraction decomposition of the integral on the left-hand side. In 1997 neither *Maple* nor *Mathematica* could evaluate (3). Today both systems can.

3.1   Binary digits of log 2

The BBP formula (3) was not discovered by a conventional analytic derivation. Instead, it was discovered via a computer-based search using the PSLQ *integer relation detection algorithm* (see Section 3.2) of mathematician-sculptor Helaman Ferguson [4], in a process that some have described as an exercise in "reverse mathematical engineering." The motivation for this search was the earlier observation by the authors of [3] that log 2 also has this arbitrary position digit calculating property. This can be seen by analyzing the classical formula

$$\log 2 \;=\; \sum_{k=1}^{\infty} \frac{1}{k2^k}, \tag{7}$$

which has been known at least since the time of Euler, and which is closely related to the functional equation for the dilogarithm.

Let $r \bmod 1$ denote the fractional part of a nonnegative real number r, and let d be a nonnegative integer. Then the binary fraction of log 2 after the "decimal" point has been shifted to the right d places can be written as

$$
\begin{aligned}
(2^d \log 2) \bmod 1 \;&=\; \left(\sum_{k=1}^{d} \frac{2^{d-k}}{k} \bmod 1 + \sum_{k=d+1}^{\infty} \frac{2^{d-k}}{k} \bmod 1 \right) \bmod 1 \\
&=\; \left(\sum_{k=1}^{d} \frac{2^{d-k} \bmod k}{k} \bmod 1 + \sum_{k=d+1}^{\infty} \frac{2^{d-k}}{k} \bmod 1 \right) \bmod 1, \tag{8}
\end{aligned}
$$

where "mod k" has been inserted in the numerator of first term since we are only interested in the fractional part of the result after division.

The operation $2^{d-k} \bmod k$ can be performed very rapidly by means of the *binary algorithm for exponentiation*. This scheme is the observation that an exponentiation operation such as 3^{17} can be performed in only five multiplications, instead of 16, by writing it as $3^{17} = ((((3^2)^2)^2)^2) \cdot 3$. Additional savings can be realized by reducing all of the intermediate multiplication results modulo k at each step. This algorithm, together with the division and summation operations indicated in the first term, can be performed in ordinary double-precision floating-point arithmetic, or, for very large calculations by using quad- or oct-precision arithmetic.

Expressing the final fractional value in binary notation yields a string of digits corresponding to the binary digits of log 2 beginning immediately after the first d digits of log 2. Computed results can be easily checked by performing this operation for two slightly different positions, say $d - 1$ and d, then checking to see that resulting digit strings properly overlap.

3.2 Hunt for the pi formula

In the wake of finding the above scheme for the binary digits of $\log 2$, the authors of [3] immediately wondered if there was a similar formula for π (none was known at the time). Their approach was to collect a list of mathematical constants (α_i) for which formulas similar in structure to the formula for $\log 2$ were known in the literature, and then to determine, by means of a PSLQ *integer relation computation*, if a nontrivial linear relation exists of the form

$$a_0 \pi + a_1 \alpha_1 + a_2 \alpha_2 + \cdots + a_n \alpha_n = 0, \tag{9}$$

where a_i are integers (because such a relation could then be solved for π to yield the desired formula). After several months of false starts, the following relation was discovered:

$$\pi = 4 \cdot {}_2F_1 \left(\begin{matrix} 1, \frac{1}{4} \\ \frac{5}{4} \end{matrix} \middle| -\frac{1}{4} \right) + 2 \arctan \left(\frac{1}{2} \right) - \log 5, \tag{10}$$

where the first term is a Gauss hypergeometric function evaluation. After writing this formula explicitly in terms of summations, the BBP formula for π was uncovered:

$$\pi = \sum_{k=0}^{\infty} \frac{1}{16^k} \left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right). \tag{11}$$

3.3 The BBP formula in action

Variants of the BBP formula have been used in numerous computations of high-index digits of π. In 1998 Colin Percival, then a 17-year-old undergraduate at Simon Fraser University in Canada, computed binary digits beginning at position one quadrillion (10^{15}). At the time, this was one of the largest, if not the largest, distributed computations ever done. More recently, in July 2010, Tsz-Wo Sze of *Yahoo! Cloud Computing*, in roughly 500 CPU-years of computing on *Apache Hadoop* clusters, found that the base-16 digits of π beginning at position 5×10^{14} (corresponding to binary position two quadrillion) are:

    0 E6C1294A ED40403F 56D2D764 026265BC A98511D0 FCFFAA10 F4D28B1B B5392B8.

The BBP formulas have also been used to confirm other computations of π. For example, in August 2010, Shigeru Kondo (a hardware engineer) and Alexander Yee (an undergraduate software engineer) computed five trillion decimal digits of π on a home-built \$18,000 machine. They found that the last 30 digits leading up to position five trillion are

    7497120374 4023826421 9484283852.

Kondo and Yee (see photos in Figure 4) used the following Chudnovsky-Ramanujan series:

$$\frac{1}{\pi} = 12 \sum_{k=0}^{\infty} \frac{(-1)^k (6k)!(13591409 + 545140134k)}{(3k)! (k!)^3 640320^{3k+3/2}}, \tag{12}$$

They did not merely evaluate this formula as written, but instead employed a clever quasi-symbolic scheme that mostly avoids the need for full-precision arithmetic.

The existence of such a rational series relies on the fact that $\mathbb{Q}(\sqrt{-163})$ has unique factorization. The sum (12) was prefigured by Ramanujan's series in 1914:

$$\frac{1}{\pi} = \frac{2\sqrt{2}}{9801} \sum_{k=0}^{\infty} \frac{(4k)! (1103 + 26390k)}{(k!)^4 396^{4k}}, \tag{13}$$
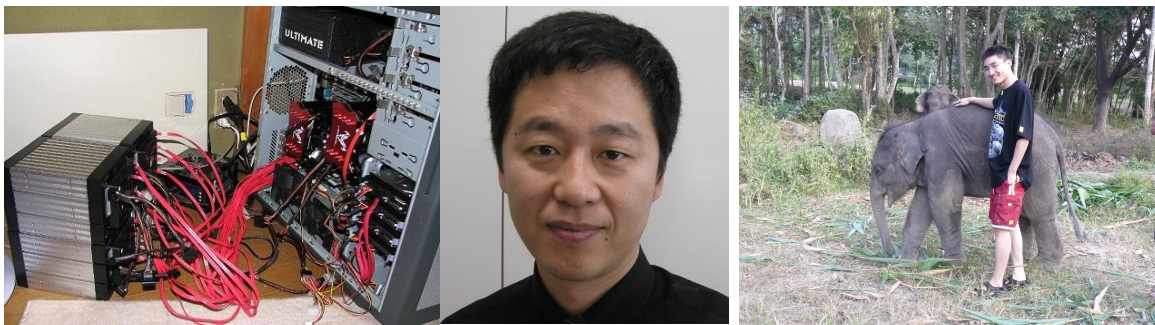
8

Figure 4: (L) Shigeru Kondo and his π-computer. (R) Alex Yee and his elephant
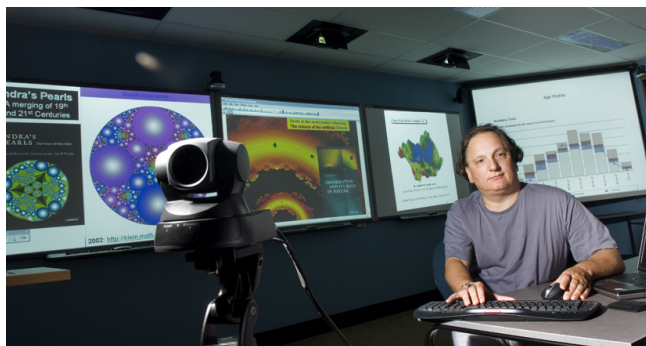

Figure 5: Jonathan Borwein in his laboratory

which lies in $\mathbb{Q}(\sqrt{58})$. This formula was used in Gosper's 1985 computation of π. Gosper's calculation also completed the first proof of (13), since only the coefficient 1103 was unproven and any other algebraic number would not have yielded over three million correct digits.

Kondo and Yee first computed their result in hexadecimal (base-16) digits. Then, in a crucial verification step, they checked hex digits near the end against the same string of digits computed using the BBP formula for π. When this test passed, they converted their entire result to decimal. The entire computation took 90 days, including 64 hours for the BBP confirmation and 8 days for base conversion to decimal. Note that the much lower time for the BBP confirmation, relative to the other two parts, greatly reduced the overall computational cost. A very detailed description of their work is available at [14].

4   Other BBP-type formulas

One question that immediately arose in the wake of the discovery of the BBP formula for π was whether there are formulas of this type for π in other number bases — in other words, formulas where the 16 in the BBP formula is replaced by some other integer, such as 3 or 10. These computer searches were largely laid to rest in 2004, when one of the present authors (Jonathan Borwein), together with Will Galway and David Borwein (Jonathan's father) showed that there are no degree-1 BBP-type formulas of *Machin-type* for π, except those whose base is a power of two [9, Thm. 3.6]. A photo of Jonathan Borwein in his laboratory is shown in Figure 5.

In the years since 1997, computer searches using the *PSLQ algorithm*, as well as conventional

9

analytic investigations, have uncovered BBP-type formulas for numerous other mathematical constants, including $\pi^2, \log^2 2, \pi \log 2, \zeta(3), \pi^3, \log^3 2, \pi^2 \log 2, \pi^4, \zeta(5)$ and Catalan's constant. BBP formulas are also known for many arctangents, and for $\log k$, $2 \le k \le 22$, although none is known for $\log 23$. These formulas and many others, together with references, are given in an online compendium [1].

One particularly intriguing fact is that whereas only binary formulas exist for π, there are both binary and ternary (base-3) formulas for π^2:

$$\pi^2 = \frac{9}{8} \sum_{k=0}^{\infty} \frac{1}{64^k} \left(\frac{16}{(6k+1)^2} - \frac{24}{(6k+2)^2} - \frac{8}{(6k+3)^2} - \frac{6}{(6k+4)^2} + \frac{1}{(6k+5)^2} \right).$$
(14)

$$\pi^2 = \frac{2}{27} \sum_{k=0}^{\infty} \frac{1}{729^k} \left(\frac{243}{(12k+1)^2} - \frac{405}{(12k+2)^2} - \frac{81}{(12k+4)^2} - \frac{27}{(12k+5)^2} \right.$$
$$\left. - \frac{72}{(12k+6)^2} - \frac{9}{(12k+7)^2} - \frac{9}{(12k+8)^2} - \frac{5}{(12k+10)^2} + \frac{1}{(12k+11)^2} \right).$$
(15)

Formula (14) appeared in [3], while formula (15) is due to Broadhurst. There are known binary BBP formulas for both $\zeta(3)$ and π^3, but no one has found a ternary formula for either.

4.1 Catalan's constant

One other mathematical constant of central interest is Eugéne Charles *Catalan's* (1814-1894) constant

$$G = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} = 0.91596559417722\ldots,$$
(16)

which is arguably the most basic constant whose irrationality and transcendence (though strongly suspected) remain unproven. Note the close connection to this formula for π^2:

$$\frac{\pi^2}{8} = \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = 1.2337005501362\ldots.$$
(17)

Formulas (16) and (17) can be viewed as the simplest Dirichlet L-series values at 2, hence our decision to use these constants in our computations.

Catalan's constant has already been the subject of some large computations. In 2009, Alexander Yee and Raymond Chan calculated G to 31.026 billion digits [13]. This computation employed two formulas, including this formula due to Ramanujan:

$$G = \frac{3}{8} \sum_{n=0}^{\infty} \frac{1}{\binom{2n}{n}(2n+1)^2} + \frac{\pi}{8} \log(2 + \sqrt{3}),$$
(18)

which can be derived from the fact that $G = -T(\pi/4) = -3/2 \cdot T(\pi/12)$, where $T(\theta) := \int_0^\theta \log \tan \sigma \, d\sigma$.

Figure 6: Berkeley bus banner with David Bailey's question about randomness of π

The BBP compendium lists two BBP-type formulas for G. The first was discovered numerically by Bailey, but both it and the second formula were subsequently proven by Kunle Adegoke, based in part on some results of Broadhurst.

For the present study, we sought a formula for G with as few terms as possible, because run time for computing with a BBP-type formula increases roughly linearly with the number of nonzero coefficients. The two formulas in the compendium have 22 and 18 nonzero coefficients. So we explored the linear space of formulas for G spanned by these two formulas, together with two known "zero relations" (BBP-type formulas whose sum is zero). This led to the following formula, which has only 16 nonzero coefficients, and was used in our computations:

$$
\begin{aligned}
G \; = \; & \frac{1}{4096} \sum_{k=0}^{\infty} \frac{1}{4096^k} \left(\frac{36864}{(24k+2)^2} - \frac{30720}{(24k+3)^2} - \frac{30720}{(24k+4)^2} - \frac{6144}{(24k+6)^2} - \frac{1536}{(24k+7)^2} \right. \\
& + \frac{2304}{(24k+9)^2} + \frac{2304}{(24k+10)^2} + \frac{768}{(24k+14)^2} + \frac{480}{(24k+15)^2} + \frac{384}{(24k+11)^2} + \frac{1536}{(24k+12)^2} \\
& \left. + \frac{24}{(24k+19)^2} - \frac{120}{(24k+20)^2} - \frac{36}{(24k+21)^2} + \frac{48}{(24k+22)^2} - \frac{6}{(24k+23)^2} \right).
\end{aligned}
\tag{19}
$$

5 BBP formulas and normality

One motivation in computing and analyzing digits of π and related constants is to explore the age-old question of whether and why these digits appear "random." Numerous computer-based statistical checks of the digits of π — unlike those of e — so far have failed to disclose any deviation from reasonable statistical norms. See, for instance, Table 4, which presents the counts of individual hexadecimal digits among the first trillion hex digits, as obtained by Yasumasa Kanada. The randomness issue inspired a banner that adorned a shuttle bus in Berkeley, California for several years — see Figure 6.

Given some positive integer b, a real number α is said to be b-normal if every m-long string of base-b digits appears in the base-b expansion of α with precisely the expected limiting frequency $1/b^m$. It follows from basic probability theory that almost all real numbers are b-normal for any specific base b and even for all bases simultaneously. But proving normality for specific constants of interest in mathematics has proven remarkably difficult.

Interest in BBP-type formulas was heightened by the 2001 observation, by one of the present authors (Bailey) and Richard Crandall, that the normality of BBP-type constants such as

$\pi, \pi^2, \log 2$ and G can be reduced to a certain hypothesis regarding the behavior of a class of chaotic iterations [5]. No proof of that general hypothesis was offered in their paper (and any proof is likely to be quite difficult), but any specific instances of this result would be quite interesting. For example, if it could be established that the iteration given by $w_0 = 0$, and

$$w_n = \left(2w_{n-1} + \frac{1}{n}\right) \bmod 1 \tag{20}$$

is equidistributed in $[0, 1)$ (i.e., is a "good" pseudorandom number generator), then, according to the Bailey-Crandall result, it would follow that $\log 2$ is 2-normal. In a similar vein, if it could be established that the iteration given by $x_0 = 0$ and

$$x_n = \left(16x_{n-1} + \frac{120n^2 - 89n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21}\right) \bmod 1 \tag{21}$$

is equidistributed in $[0, 1)$, then it would follow that π is 2-normal.

Giving further hope to these studies is the recent extension of these methods, also by Bailey and Crandall, to a rigorous proof of normality for an uncountably infinite class of real numbers. Given a real number r in $[0, 1)$, let r_k denote the k-th binary digit of r. Then the real number

$$\alpha_{2,3}(r) = \sum_{k=0}^{\infty} \frac{1}{3^k 2^{3^k + r_k}} \tag{22}$$

is 2-normal. For example, the constant $\alpha_{2,3}(0) = \sum_{k \geq 0} 1/(3^k 2^{3^k}) = 1.0434782608695652217\ldots$ is provably 2-normal. A similar result applies if 2 and 3 in this formula are replaced by any pair of co-prime integers (b, c) greater than one [6, 7, 9].

5.1 A curious hexadecimal conjecture

It is tantalizing that if, using (21), one calculates the hexadecimal digit sequence

$$y_n = \lfloor 16x_n \rfloor \tag{23}$$

(where $\lfloor \cdot \rfloor$ denotes greatest integer), then the sequence (y_n) appears to perfectly (not just approximately) produce the hexadecimal expansion of π. In explicit computations, we checked that the first 10,000,000 hexadecimal digits generated by this sequence are *identical* with the first 10,000,000 hexadecimal digits of $\pi - 3$. This is a fairly difficult computation, as it requires roughly n^2 bit-operations, and is not easily performed on a parallel computer system. In our implementation, computing $2,000,000$ hex digits with (23), using *Maple*, required 17.3 hours on a laptop. Computing 4,100,000 using *Mathematica*, with a more refined implementation, required 46.5 hours. The full confirmation, using a C++ program, took 433,192 seconds (120.3 hours) on a IBM Power 780 (model: 9179-MHB, clock speed: 3.864 GHz). All these outputs were confirmed against stored hex digits of π in the software section of http://www.experimentalmath.info.

Conjecture 1 *The sequence $\lfloor 16x_n \rfloor$, where (x_n) is the sequence of iterates defined in equation (21), generates precisely the hexadecimal expansion of $\pi - 3$.*

Hex Digit	Occurrences
0	62499881108
1	62500212206
2	62499924780
3	62500188844
4	62499807368
5	62500007205
6	62499925426
7	62499878794
8	62500216752
9	62500120671
A	62500266095
B	62499955595
C	62500188610
D	62499613666
E	62499875079
F	62499937801
Total	1000000000000

Table 4: Digit counts in the first trillion hexadecimal (base-16) digits of π. Note that deviations from the average value 62,500,000,000 occur only after the first six digits, as expected.

We can learn more. Let $||x - y|| = \min(|x - y|, |1 - (x - y)|)$ denote the "wrapped" distance between reals x and y in $[0, 1)$. The base-16 expansion of π, which we denote π_n, satisfies

$$||\pi_n - x_n|| \leq \sum_{k=n+1}^{\infty} \frac{120k^2 - 89k + 16}{16^{k-n}(512k^4 - 1024k^3 + 712k^2 - 206k + 21)} \approx \frac{1}{64(n+1)^2}, \qquad (24)$$

so that, upon summing from some N to infinity, we obtain the finite value

$$\sum_{n=N}^{\infty} ||\pi_n - x_n|| \leq \frac{1}{64(N+1)}. \qquad (25)$$

Heuristically, let us assume that the π_n are independent, uniformly distributed random variables in $(0, 1)$, and let $\delta_n = ||\alpha_n - x_n||$. Note that an error (i.e., an instance where x_n lies in a different subinterval of the unit interval than π_n, so that the corresponding hex digits don't match) can only occur when π_n is within δ_n of one of the points $(0, 1/16, 2/16, \cdots, 15/16)$. Since $x_n < \pi_n$ for all n (where $<$ is interpreted in the wrapped sense when x_n is slightly less than one), this event has probability $16\delta_n$. Then the fact that the sum (25) has a finite value implies, by the first *Borel-Cantelli* lemma, that there can only be finitely many errors. Further, the small value of the sum (25), even when $N = 1$, suggests that it is unlikely that any errors will be observed. If we set $N = 10,000,001$ in (25), since we know there are no errors in the first 10,000,000 elements, then we obtain an upper bound of 1.563×10^{-9} which suggests it is truly unlikely that errors will ever occur.

A similar correspondence can be seen between iterates of (20) and the binary digits of $\log 2$. In particular, let $z_n = \lfloor 2w_n \rfloor$, where w_n is given in (20). Then since the sum of the error terms

Figure 7: Andrew Mattingly, Blue Gene/P, and Glenn Wightwick

for $\log 2$, corresponding to (25), is infinite, it follows by the second Borel-Cantelli lemma that discrepancies between (z_n) and the binary digits of $\log 2$ can be expected to appear indefinitely, but with decreasing frequency. Indeed, in computations that we have done, we have found that the sequence (z_n) disagrees with 10 of the first 20 binary digits of $\log 2$, but in only one position over the range 5000 to 8000.

6   Computing digits of π^2 and Catalan's constant

In illustration of this theory, we now present the results of computations of high-index binary digits of π^2, ternary digits of π^2, and binary digits of Catalan's constant, based on formulas (14), (15) and (19), respectively. These calculations were performed on a 4-rack *BlueGene/P* system at IBM's Benchmarking Centre in Rochester, Minnesota, USA. This is a shared facility, so calculations were conducted over a several month period, where, at any given time, none, some or all of the system was available. It was programmed remotely from Australia, which permitted the system to be used off-hours. Sometimes it helps to be in a different time zone!

1. *Base-64 digits of π^2 beginning at position 10 trillion.* The first run, which produced base-64 digits starting from position $10^{12} - 1$, required an average of 253,529 seconds per thread, and was subdivided into seven partitions of 2048 threads each, so the total cost was $7 \cdot 2048 \cdot 253529 = 3.6 \times 10^9$ CPU-seconds. Each rack of the IBM system features 4096 cores, so the total cost is 10.3 "rack-days." The second run, which produced base-64 digits starting from position 10^{12}, completed in nearly the same run time (within a few minutes). The two resulting base-8 digit strings are

   ```
   75|601145053032364757245000057432627545303630524163506 34|573227604
     |601145053032364757245000057432627545303630524163506 34|220210566
   ```

   (each pair of base-8 digits corresponds to a base-64 digit). Here the digits in agreement are delimited by |. Note that 53 consecutive base-8 digits (or, equivalently, 159 consecutive binary digits) are in perfect agreement.

2. *Base-729 digits of π^2 beginning at position 10 trillion.* In this case, the two runs each required an average of 795,773 seconds per thread, similarly subdivided as above, so that the total cost was 6.5×10^9 CPU-seconds, or 18.4 "rack-days." The two resulting base-9 digit strings are

14

CONSTANT	n'	d	#ITERS $(\times 10^{15})$	TIME/ITER (microsec)	TIME (yr)	WITH VERIFY	TOTAL (yr)	O'HEAD (%)	FLOPS $(\times 10^{18})$
π^2 base-2^6	5	10^{13}	2.16	1.424	97.43	194.87	230.35	18.2	2.58
π^2 base-3^6	9	10^{13}	3.89	1.424	175.38	350.76	413.16	17.8	4.65
G base-4^6	16	10^{13}	6.91	1.424	311.79	623.58	735.02	17.9	8.26

Table 5: The scale of our computations. We estimate 4.5 quad-double operations per iteration and, , that each costs 266 single-precision operations. The total cost in single-precision operations is given in the last column. This total includes overhead which is largely due to a rounding operation that we implemented using bit-masking.

```
001|1226448506454858317711113521016285604832345 3468|10565567635862
   |1226448506454858317711113521016285604832345 3468|04744867134524
```

(each triplet of base-9 digits corresponds to one base-729 digit). Note here that 47 consecutive base-9 digits (94 consecutive base-3 digits) are in perfect agreement.

3. *Base-4096 digits of Catalan's constant beginning at position 10 trillion.* These two runs each required 707,857 seconds per thread, but in this case were subdivided into eight partitions of 2048 threads each, so that the total cost was 1.2×10^{10} CPU-seconds, or 32.8 "rack-days." The two resulting base-8 digit strings are

```
0176|347050537747770511226133716201252573272173 24522|6000177545727
    |347050537747770511226133716201252573272173 24522|5703510516602
```

(each quadruplet of base-8 digits corresponds to one base-4096 digit). Note that 47 consecutive base-8 digits (141 consecutive binary digits) are in perfect agreement.

These long strings of consecutively agreeing digits, beginning with the target digit, provide a compelling level of statistical confidence in the results. In the first case, for instance, note that the probability that 32 pairs of randomly chosen base-8 digits are in perfect agreement is roughly 1.2×10^{-29}. Even if one discards, say, the final six base-8 digits as a 1-in-262,144 statistical safeguard against numerical round-off error, one would still have 24 consecutive base-8 digits in perfect agreement, with a corresponding probability of 2.1×10^{-22}. Now strictly speaking, one cannot define a valid probability measure on digits of π^2, but nonetheless, from a practical point of view, such analysis provides a very high level of statistical confidence that the results have been correctly computed.

For this reason, computations of π and the like are a favorite tool for the integrity testing for computer system hardware and software. If either run of a paired computation of π succumbs to even a single fault in the course of the computation, then typically the final results will disagree almost completely. For example, in 1986, a similar pair of computations of π disclosed some subtle but substantial hardware errors in an early model of the Cray-2 supercomputer. Indeed, the calculations we have done arguably constitute the most strenuous integrity test ever performed on the BlueGene/P system. Table 5 gives some sense of the scale of the three record computations, which used more than 135 "rack-days," 1378 serial CPU-years and more than

Digit	0	1	2	3	4	5	6	7
base-2 (141)	0.454	0.546	-	-	-	-	-	-
base-4 (70)	0.171	0.329	0.229	0.271	-	-	-	-
base-8 (47)	0.085	0.128	0.213	0.128	0.064	0.128	0.043	0.213

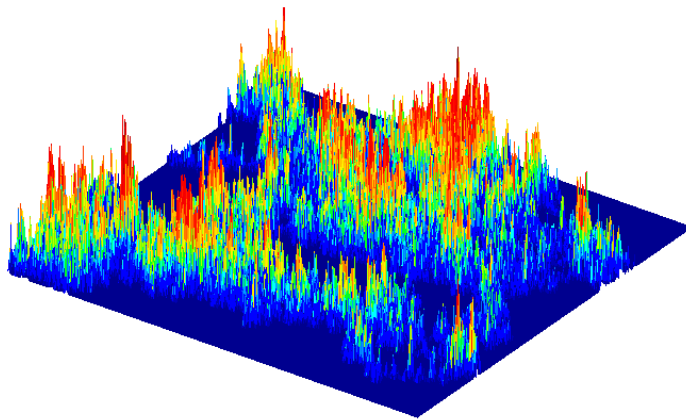Table 6: Base-4096 digits of G beginning at position 10 trillion: digit proportions

Figure 8: A random walk on a million digits of Catalan's constant

1.549×10^{19} floating point operations. This is comparable to the cost of the most sophisticated animated movies as of the present time (2011).

For the sake of completeness, in Table 5 we also record the one, two and three-bit frequency counts from our Catalan computation.

7 Future directions

It's a clue.
A never repeating or ending chain, the total timeless catalogue,
elusive sequences, sum of the universe.
This riddle of nature begs:
Can the totality see no pattern, revealing order as reality's disguise?[1]

It is ironic that in an age when even pillars such as Fermat's Last Theorem and the Poincairé conjecture have succumbed to the brilliance of modern mathematics, that one of the most elementary mathematical hypotheses, namely whether (and why) the digits of π or other constants, such as $\log 2, \pi^2$ or G (see Figure 8), are "random," remains unanswered. In particular, proving that π (or $\log 2, \pi^2$ or G) is b-normal in some integer base b remains frustratingly elusive. Even much weaker results, for instance the simple assertion that a one appears in the binary expansion of π (or $\log 2, \pi^2$ or G) with limiting frequency 1/2 (which assertion has been amply

---

[1] A self-referent digit-mnemonic for pi from `http://www.newscientist.com/blogs/culturelab/2010/03/happy-pi-day.php`.

affirmed in numerous computations over the years), remain unproven and largely inaccessible at the present time.

Almost as much ignorance extends to simple algebraic irrationals such as $\sqrt{2}$. In this case it is now known that the number of ones in the first n binary digits of $\sqrt{2}$ must be at least of the order of $\sqrt{n}$, with similar results for other algebraic irrationals [2]. But this is a very weak result, given that this limiting ratio is almost certainly $1/2$, not only for $\sqrt{2}$ but more generally for all algebraic irrationals.

Nor can we prove much about continued fractions for various constants, except for a few well-known results for special cases such as quadratic irrationals, ratios of Bessel functions, and certain expressions involving exponential functions.

For these reasons, there is continuing interest in the theory of BBP-type constants, since, as mentioned, there is an intriguing connection between BBP-type formulas and certain chaotic iterations that are akin to pseudorandom number generators. If these connections can be strengthened, then perhaps normality proofs could be obtained for a wide range of polylogarithmic constants, possibly including $\pi, \log 2, \pi^2$ and G.

As settings change, so do questions. Until the question of efficient single-digit extraction was asked, our ignorance about such issues was not exposed. The case of the exponential series

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{26}$$

is illustrative. For present purposes, the convergence rate in (26) is too good.

Conjecture 2 *There is no BBP formula for e. Moreover, there is no way to extract individual digits of e significantly more rapidly than by computing the first n digits.*

The same could be conjectured about other numbers including the *Euler-Mascheroni constant* $\gamma := 0.57721566490153\ldots$. In short, until vastly stronger mathematical results are obtained in this area, there will doubtless be continuing interest in computing digits of these constants. In the present vacuum, that is perhaps all that we can do.

References

[1] David H. Bailey, "BBP-Type Formulas," manuscript, 2011, available at `http://crd.lbl.gov/~dhbailey/dhbpapers/bbp-formulas.pdf`. Web data-base version at `http://www.bbp.carma.newcastle.edu.au`.

[2] David H. Bailey, Jonathan M. Borwein, Richard E. Crandall and Carl Pomerance, "On the Binary Expansions of Algebraic Numbers," *Journal of Number Theory Bordeaux*, vol. 16 (2004), pg. 487–518.

[3] David H. Bailey, Peter B. Borwein and Simon Plouffe, "On The Rapid Computation of Various Polylogarithmic Constants," *Mathematics of Computation*, vol. 66, no. 218, 1997, pg. 903–913.

[4] David H. Bailey and David J. Broadhurst, "Parallel Integer Relation Detection: Techniques and Applications," *Mathematics of Computation*, vol. 70, no. 236 (Oct 2000), pg. 1719–1736.

[5] David H. Bailey and Richard E. Crandall, "On the Random Character of Fundamental Constant Expansions," *Experimental Mathematics*, vol. 10, no. 2 (Jun 2001), pg. 175–190.

[6] David H. Bailey and Richard E. Crandall, "Random Generators and Normal Numbers," *Experimental Mathematics*, vol. 11, no. 4 (2002), pg. 527–546;

[7] David H. Bailey and Michal Misiurewicz, "A Strong Hot Spot Theorem," *Proceedings of the American Mathematical Society*, vol. 134 (2006), no. 9, pg. 2495–2501.

[8] J.M. Borwein, "The Life of Pi". Extended and updated version of "La vita di pi greco," volume 2 of *Mathematics and Culture, La matematica: Problemi e teoremi*, Guilio Einaudi Editori, Turino, Italian, 2008) (French, in Press) *Berggren Festschrift Volume*. Submitted to *A Handbook of the Ancient and Medieval Mathematical Sciences in Honour of J. L. Berggren*, Springer. May 2010.

[9] J.M. Borwein and D.H. Bailey, *Mathematics by Experiment: Plausible Reasoning in the 21st Century* A.K. Peters Ltd, 2004. Expanded Second Edition, 2008.

[10] Jonathan M. Borwein and Peter B. Borwein, *Pi and the AGM*, John Wiley and Sons, New York, 1987.

[11] L. Berggren, J.M. Borwein and P.B. Borwein, *Pi: a Source Book,* Springer-Verlag, (1997), (2000), (2004). Fourth edition in preparation, 2011.

[12] Jonathan M. Borwein, William F. Galway and David Borwein, "Finding and Excluding b-ary Machin-Type BBP Formulae," *Canadian Journal of Mathematics*, vol. 56 (2004), pg. 1339–1342.

[13] Alexander Yee, "Large Computations," 7 Mar 2011, available at `http://www.numberworld.org/nagisa_runs/computations.html`.

[14] Alexander Yee and Shigeru Kondo, "5 Trillion Digits of Pi — New World Record," 7 Mar 2011, available at `http://www.numberworld.org/misc_runs/pi-5t/details.html`.